1 KEKER & VAN NEST, LLP
   DARALYN J. DURIE - #169825
2 EUGENE M. PAIGE - #202849
   RYAN M. KENT - #220441
3 SONALI D. MAITRA – #254896
   710 Sansome Street
4 San Francisco, CA 94111-1704
   Telephone: (415) 391-5400
5 Facsimile: (415) 397-7188

6 Attorneys for Defendants
   WELLS FARGO BANK, N.A. and
7 WELLS FARGO FUNDS MANAGEMENT, LLC

8

9   UNITED STATES DISTRICT COURT

10   NORTHERN DISTRICT OF CALIFORNIA

11   SAN FRANCISCO DIVISION

12

13	PHOENIX SOLUTIONS, INC., a California corporation,	Case No. CV 08-0863 MHP
14		**WELLS FARGO'S RESPONSIVE MARKMAN BRIEF**
	Plaintiff,	
15		Date: December 11, 2008
16	v.	Time: 2:30 p.m.
		Dept: Courtroom 15, 18th floor
17	WELLS FARGO BANK, N.A., a Delaware corporation, and WELLS FARGO FUNDS	Judge: Hon. Marilyn Hall Patel
18	MANAGEMENT, LLC, a Delaware limited liability company,	Date Comp. Filed: February 8, 2008
19	Defendants.	Trial Date: TBD

20

21

22

23

24

25

26

27

28

TABLE OF CONTENTS

i

430233.02

TABLE OF CONTENTS
(cont'd)

ii

430233.02

TABLE OF AUTHORITIES

WELLS FARGO'S RESPONSIVE MARKMAN BRIEF
CASE NO. CV 08-0863 MHP

430233.02

I.   PRELIMINARY STATEMENT

"It is axiomatic that claims are construed the same way for both invalidity and infringement." *Amgen, Inc. v. Hoechst Marion Roussel, Inc.*, 314 F.3d 1313, 1330 (Fed. Cir. 2003). Phoenix seeks to violate this fundamental axiom: it tries to fend off Wells Fargo's motion for summary judgment of invalidity by insisting that the prior art must have certain attributes in order to anticipate its claims, but those supposedly necessary elements are nowhere to be found in its proposed claim constructions. Phoenix certainly has the right to argue for a broad construction of its claims. But it cannot do so while at the same time arguing that prior art that would fall within those broad constructions is not anticipatory. For example, Phoenix has argued in opposing summary judgment that it is not sufficient to show that the "pruning" of a manually written grammar can be allocated between a client and a server, Phoenix Solutions, Inc.'s Opposition to Wells Fargo's Motion for Summary Judgment, Docket No. 86 ("SJ Opp.") at 14:10-19; the item connected to the server in a "remote speech capturing system" is a "client" and not just a "computing system," SJ Opp. at 19:4-6; and a natural language engine meets the claim limitations only when it is able automatically to fill in gaps in grammars in order to accommodate speech input that varies from the expected speech input, SJ Opp. at 21:16-22:7. Yet these supposedly essential features have not made their way into Phoenix's proposed constructions or the arguments in its *Markman* briefing.

Nor is that the only internal inconsistency in Phoenix's approach to claim construction. While Phoenix pays lip service to another axiom – that limitations should not be imported into the claims from the specification – Phoenix proceeds to ignore it. Phoenix argues that "a communications channel" requires that data be sent "with an Internet based protocol," but nothing in the claim language or the specification requires that the ordinary meaning of the term "communications channel" be restricted in this way. At the same time, Phoenix castigates Wells Fargo for looking to the specification to understand the meaning of claim terms like "morphological analysis" and "phrase analysis" that are explicitly discussed therein. But, as Phoenix itself concedes, the specification "is always highly relevant to the claim construction analysis. Usually, it is dispositive; it is the single best guide to the meaning of a disputed term."

1

1   *Phillips v. AWH Corp.*, 415 F.3d 1303, 1315 (Fed. Cir. 2005). Here, a review of the

2   specification helps to elucidate the invention that Phoenix sought to claim.

II.   FACTUAL BACKGROUND

4   Phoenix's Markman brief suggests that Phoenix invented "a natural language interactive

5   voice response ("IVR") system that is superior to a conventional touch tone system." Phoenix

6   Solutions Inc.'s Opening Claim Construction Brief, Docket No. 116, ("Br.") at 1:17-18. From

7   this, one might think that Phoenix had invented the speech-enabled IVR – the type of system that

8   allows a user to call a customer hotline and obtain information about stock prices or bank

9   accounts. However, as is discussed in more detail in Wells Fargo's summary judgment motion,

10   speech-enabled IVRs long-predated Phoenix's alleged invention. Thus, at most, Phoenix sought

11   to patent specific improvements to such a speech-enabled IVR.

**1.   The prior art taught speech recognition, natural language processing and
distributed computing**

In the common patent specification, Phoenix does not contend that it invented any of the

speech recognition and understanding technology underlying its claims. The patent notes that

the prior art taught the use of speech vectors such as MFCC vectors and Hidden Markov Models

(HMM) to recognize speech, the theory of which was published in "classic papers" in the 1960s

and 1970s, including two papers authored by Dr. Lawrence Rabiner, Wells Fargo's expert in this

case. Declaration of R. Joseph Trojan in Support of Plaintiff's Opening Claim Construction

Brief ("Trojan Decl."), Ex. 1 at 3:34-67 ('846 patent). The prior art also taught "a distributed

approach to speech recognition," including the implementation of "an HMM-based distributed

speech recognition system" and "the notion of tailoring a recognition process to a set of available

computational resources." *Id*. at 5:1-62. The patent explains that prior art research had

combined HMMs with neural networks to permit continuous speech recognition, resulting in a

paper by Morgan et al. on "Hybrid Neural Network/Hidden Markov Model Systems for

Continuous Speech Recognition." *Id*. at 4:5-10. Existing commercial programs such as one

offered by IBM could recognize speech and even interact with websites to allow "user control of

the interface (opening, closing files) and searching (by using previously-trained URLs)." *Id*. at

2

2:15-17. Finally, the patent notes the use in the prior art of Natural Language Processing, which

"is concerned with the parsing, understanding and indexing of transcribed utterances and larger

linguistic units." *Id*. at 4:26-29.

2.   Phoenix develops a system that allocates the generation of speech vectors between a client and a server

So what did Phoenix purport to add to the art? While the specification does not provide a

crystal clear answer, it does provide some clues. The patent specification sets forth the invention

as embodied in the Natural Language Query System (NLQS), an acronym used by inventor Ian

Bennett for the system that he purports to have developed.[1]

Fig. 1
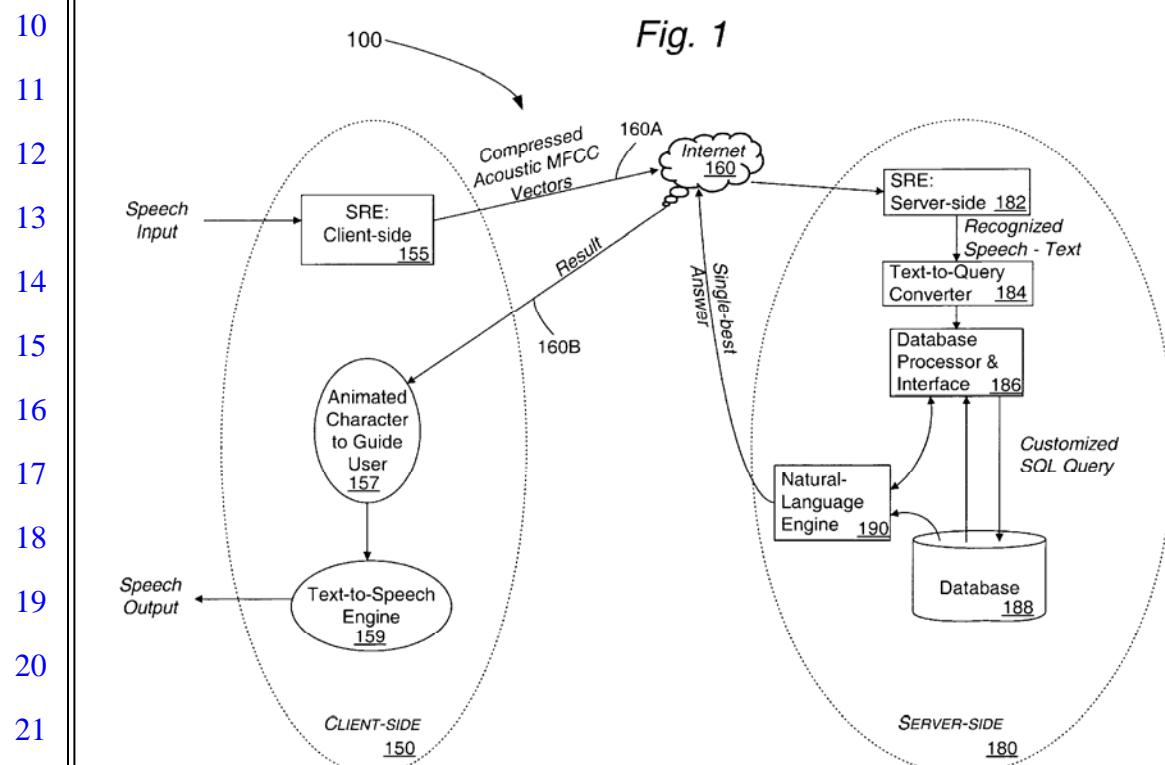
Trojan Decl., Ex. 1 at Figure 1.

The specification describes a key feature of the invention, as distinguished from the prior

art, to be dividing certain speech recognition operations between a client (150) and a server

---

[1] Mr. Bennett claims to have developed a working prototype of the NLQS, but Phoenix has refused to make that system available for inspection in discovery, so the extent to which Phoenix ever had a working embodiment of its "invention" remains to be seen. Declaration of Eugene M. Paige in Support of Wells Fargo's Responsive Markman Brief ("Paige Decl."), Ex. 1. Bennett's initial schematic for the NLQS was nothing more than a schematic drawing showing a variety of

WELLS FARGO'S RESPONSIVE MARKMAN BRIEF
CASE NO. CV 08-0863 MHP

1   (180). The client platform "can be as simple as a personal digital assistant or cell-phone, or as

2   sophisticated as a high-end desktop PC." Trojan Decl., Ex. 1 at 10:32-32; 10:41-43. The

3   specification explains that once the spoken question is captured at the client, "the question is

4   processed partially by NLQS client-side software resident in the client's machine. **The output**

5   **of this partial processing is a set of speech vectors** that are transported to the server …." *Id*. at

6   7:22-26. There is no suggestion in the specification that the "partially processed speech signal

7   data" sent from the client to the server could be anything other than speech vectors.

8   The specification describes the preferred speech vectors as MFCC parameters (also

9   referred to as MFCC coefficients) that are "extracted from the speech utterance continuously

10 until silence is detected." *Id*. at 22:30-32. The generation of these speech vectors "represents the

11 first phase of processing of the input speech signal." *Id*. at 22:32-33. The initial set of MFCC

12 vectors, together with an energy term, "together make up the partially processed speech data

13 transmitted in compressed form from the user's client system to the remote server side." *Id*. at

14 16:22-26.

15   The server then can optionally calculate two types of additional coefficients from the data

16 that has been sent – "the delta and acceleration coefficients representing change in each of the 13

17 [MFCC] values from frame to frame" – in order to "enhance" the performance of the recognition

18 engine. *Id*. at 16:27-32. These additional coefficients "are computed during a second partial

19 speech signal processing phase to complete the initial processing of the speech signal." *Id*. at

20 16:33-35. The specification explains that because only "capture of the speech signals, MFCC

21 vector extraction and compression are implemented on the client's machine during a first partial

22 processing phase …. the routine is thus streamlined and simple enough to be implemented within

23 a browser program … [such that] even very "thin" client platforms can be supported." *Id*. at

24 16:61-17:1. Indeed, the specification repeatedly touts the ability to split the derivation of the

25 speech vectors between the client and the server as a key feature of the invention: "As explained

26 earlier, to conserve latency time between the client and server, a smaller number of bytes (just

27 the 13 MFCC coefficients) are sent from client side system 150 to server side system 180. … So,

28 _____

publicly available components strung together. *Id.*, Ex. 2; see also *id*., Ex. 3 at 101:19-113:1.

1   the present invention can accommodate even the worst-case scenario where the client's machine

2   may be quite thin and may have just enough resources to capture the speech input data and do

3   minimal processing." Ex. 1 at 26:65-27:1; 27:15-18. In the file history, Phoenix confirmed that,

4   in this scenario involving a very "thin" client, the first set of coefficients would be sent from the

5   client to the server and the remaining acceleration and delta coefficients would be generated at

6   the server: "in some primitive client devices, it is likely that there would be no additional

7   computation made of additional acceleration/delta coefficients on the client side." Paige Decl.,

8   Ex. 4 ('846 patent Amendment A) at 19; *see also id*., Ex. 5 ('846 patent Amendment B) at 17

9   ("[I]f the invention were embodied in an MFCC application it could send the entire set of

10 acoustic features extracted from the speech utterance in some cases [from the client to the server]

11 …. and in others, it could send only a portion (i.e., perhaps omitting the delta and/or acceleration

12 coefficients.").

13   Phoenix also touted the derivation of speech vectors at both the client and the server as a

14 key distinction over the prior art. The specification concedes that the bare idea of distributed

15 voice recognition was not new: the prior art taught "an implementation of a distributed voice

16 recognition system between a telephony-based handset and a remote station." Trojan Decl., Ex.

17 1 at 5:25-31. However, in the prior art, "all of the word recognition operations seem to take

18 place" at one location (there, the client). *Id*. at 5:32-33. The specification discusses the

19 possibility that "in some environments, nonetheless, it is conceivable that the MFCC delta

20 parameters and MFCC acceleration parameters can also be computed at client side system

21 150…." Trojan Decl., Ex. 1 at 22:40-42. "Therefore, for some applications, the allocation of

22 signal processing responsibilities may be partitioned differently, to the point where in fact both

23 phases of the speech signal processing may take place at the client side system 150 so that the

24 speech signal is completely – rather than partially – processed and transmitted for conversion

25 into a query at server side system 180." *Id*. at 22:54-60. But while the specification contains this

26 explicit disclosure of deriving all the speech vectors at the client, there is no discussion of the

27 possibility of transmitting anything other than speech vectors from the client to the server and

28 thus no discussion of deriving all the vectors at the server.

430233.02

3.   A natural language engine analyzes the text output by the speech recognizer to determine the single best answer to the user's question

After the server uses the speech vectors to generate recognized text, the system tries to find the best answer to the user's question through a natural language process. A database contains a set of question and answer pairs. The invention tries to match the user's spoken question to the most relevant question in the database. First, the system identifies a set of possibly relevant questions stored in the database. Second, the user's question is compared to each of the identified questions using a natural language analysis, and the system identifies the single question that most closely matches the user's query. The patent explains that this process involves the user's question being understood "lexically, as opposed to just being 'recognized.'" Trojan Decl., Ex. 1 at Col. 12:16-17. Once the single most relevant question has been matched to the user's query, the system then returns the answer that is paired to that question as the "single best answer."

III.   ARGUMENT[2]

1.   First set of speech data values from said speech utterance signals, said first set of speech data values being insufficient by themselves for permitting recognition of words articulated in said speech utterance ('846 claims 11, 17, 20, 21)

Speech data being characterized by a data content that is substantially inadequate by itself for permitting recognition of words articulated in said query ('977 claims 1, 6, 7, 10)[3]

Wells Fargo proposed constructions:

speech utterance signals	analog or digital waveforms that represent speech
speech data values	acoustic coefficients derived from the spectral analysis of speech utterance signals

Phoenix proposed constructions:

'846 patent entire phrase	speech data is derived from the user's utterance. This data is not yet sufficiently processed to allow a speech recognition engine to identify the

---

[2] The United States District Court for the Central District of California recently issued a *Markman* ruling construing terms in other patents in the Phoenix portfolio that are not asserted in this suit, and two claim terms from the '977 patent, client platform and signal processing functions, neither of which is directly at issue here. However, for the Court's convenience, a copy of that ruling is attached as Exhibit 6 to the Paige Declaration.

[3] For ease of reference, the numbering of each part corresponds to the Joint Claim Construction Statement filed by the parties.

6

430233.02

	word spoken by the user.
'977 patent entire phrase	speech data received by the server has not been processed sufficiently into a form that allows a speech recognizer to determine individual words and phrases. Additional processing must be done (*i.e.*, such as generating acoustic features and/or a complete speech observation vector) to put it into a form suitable for use by a speech recognition engine to complete recognition.

a.   Speech Utterance Signals

Wells Fargo does not try to limit the definition of speech utterance signals by arguing that speech utterance signals must "equate to waveforms," as Phoenix suggests. Instead, Wells Fargo is simply trying to make clear that speech signals can be represented by waveforms, whether analog or digital. Phoenix does not offer any evidence or argument to the contrary. Nor can it, given that speech is comprised of acoustic waveforms. Declaration of Dr. Lawrence Rabiner ("Rabiner Decl.") at ¶ 13.

b.   Speech Data Values[4]

Phoenix does not offer any proposed construction of "speech data values." Instead, Phoenix proposes to rewrite the claims of the '846 patent to refer to "speech data," the claim term that appears in the '977 patent (and in other, unasserted, claims of the '846 patent).

"Speech data values" are data values that represent speech. The only kind of such data values referenced in the specification are acoustic coefficients, also called speech vectors, which are derived from a spectral analysis of speech utterances. The Summary of Invention explains that the output of partial processing at the client "is a set of speech vectors" that are used to recognize the speech. Trojan Decl., Ex. 1 at 7:22-26. The Detailed Description of the Invention confirms that "the output of the partial processing done by SRE (Speech Recognition Engine) 155 is a set of speech vectors….." *Id*. at 11:6-7. And the examiner had the same understanding – in the statement of reasons for allowance of the '846 patent, he noted that Phoenix had pointed out that "their client can produce first data values that are cepstral coefficient values and their

---

[4] Wells Fargo agrees that it is not inherent in the definition of "the first set of speech data values" that the first set must always be used to create a second set speech data values. Instead, that limitation is simply a function of the language of claim 11 of the '846 patent and its dependents, which specifies that the second set of speech data values are derived from the first set of speech data values.

7

1   server can produce second data values that are delta and acceleration coefficient values."[5] Paige

2   Decl., Ex. 7 ('846 patent Notice of Allowance) at 3.

3   Phoenix disagrees with Wells Fargo's proposed construction, but it does not identify any

4   other kind of data value that represents speech. Instead, Phoenix suggests (albeit inferentially)

5   that detecting silence is a "speech operation" and thus that not all "speech recognition

6   operations" require the generation of acoustic coefficients. First, there's an important distinction

7   between a "speech operation" and a "speech **recognition** operation," which Phoenix

8   conveniently glosses over. Second, a speech recognition operation is not the same thing as a

9   speech data value; one is a process, and the other is a result of a process. And, in the case of

10 detecting silence, there is no **speech** data value corresponding to silence because there is no

11 **speech** to which such a data value could relate. Unsurprisingly, then, nothing in the specification

12 suggests that "detecting silence" has anything to do with the creation of speech data values.

13 Indeed, as the Court can see from reading the passage quoted by Phoenix, the speech data values

14 were *already* derived before being sent to the server by the transmission circuit "before silence is

15 detected." Thus, the fact that detecting silence may (optionally) be involved in speech

16 recognition is beside the point. Many processes can be involved in recognizing speech, such as

17 loading dictionaries, but those are not the processes required to create speech data values.

18   Phoenix's mix-and-match approach to the claims of the '846 patent does not support its

19 contentions either. Phoenix points to claims 28 and 29, but those claims depend from (non-

20 asserted) claim 22, which does not contain the phrase "speech data values" *at all*. Instead, it

21 addresses the creation of "partially recognized speech data" rather than speech data values. This

22 Court need not determine what "partially recognized speech data" might be; it is sufficient that

23 those claims are irrelevant to the proper construction of "speech data values."

24   **c.   The claims as a whole are indefinite because they are logically incoherent**

25   <u>'977 Claim 1</u>

26   Claim 1 of the '977 patent requires that the speech data that is received by the server have

27 ─────────────────────

28 [5] The only kind of speech vectors described in the specification are MFCC vectors. However, contrary to Phoenix' argument, Wells Fargo's construction is not so limited: it covers any kind of

8

430233.02

1   a data content that is substantially inadequate by itself for permitting recognition of words

2   articulated in said speech query. But the claim also requires that a speech recognition routine

3   running on the server computing system complete recognition of the speech query using the

4   speech data and the data content. This is logically impossible. If the data content of the speech

5   data is inadequate to permit word recognition, it makes no sense to say that the system completes

6   recognition using that inadequate data content. *See* Rabiner Decl. ¶¶ 20-21. *See Chef Am., Inc.*

7   *v. Lamb-Weston, Inc.*, 358 F.3d 1371, 1373-74 (Fed. Cir. 2004) ("where as here, claims are

8   susceptible to only one reasonable interpretation and that interpretation results in a nonsensical

9   construction of the claim as a whole, the claim must be invalidated").

10   Phoenix tries to escape from this logical conundrum by arguing that the claim is not

11   addressed to the data content of the speech data but rather the extent to which the speech data has

12   been processed – that the speech data received from the client has not been sufficiently processed

13   for speech recognition.

14   *First*, Phoenix's construction simply cannot be reconciled with the unambiguous

15   language of the claim itself. Nothing in the claim language is addressed to whether the speech

16   data values (or the data content) are sufficiently "processed" for recognition. That ends the

17   inquiry, because "courts may not redraft claims…. [Courts] construe the claim as written, not as

18   the patentees wish they had written it."

19   *Second*, Phoenix is just plain wrong. Phoenix argues that a "partial" set of speech

20   vectors (such as the 12 MFCC vectors minus the delta and acceleration coefficients) is "partially

21   processed" information that is "not recognizable by the speech recognition engine." Br. 3:16-17.

22   In fact, however, those speech vectors do not need to be further "processed" before they are input

23   into a speech recognition engine. Rabiner Decl. ¶10. Indeed, the specification does not describe

24   further processing of the MFCC coefficients themselves – those coefficients remain

25   unchanged—but instead describes the derivation of **additional** coefficients from the information

26   in the original set of 12 coefficients, representing the degree to which those coefficients are

27   changing over time. Trojan Decl., Ex. 1 at 16:30-36. The original 12 MFCC coefficients can be

28

acoustic coefficients that can be used to represent speech. Rabiner Declaration ¶ 14.

9

430233.02

1   (and are) input directly into the engine that decodes speech. Rabiner Decl. ¶ 10. And the

2   specification expressly notes that the derivation of the delta and acceleration coefficients merely

3   **enhances** the speech recognition process, not that those coefficients are necessary for the speech

4   recognition process to take place. Trojan Decl., Ex. 1 at 16:27-30.

5   <u>'846 Claim 11</u>

6   Claim 11 of the '846 suffers from a similar problem. Claim 11 provides that the first set

7   of speech data values are insufficient by themselves for permitting recognition of words, that the

8   first set of speech data values are transmitted to a second signal processing circuit and that,

9   "based on receiving and processing said speech data values," the second signal processing circuit

10 generates a second set of speech data values that "contain sufficient information to be usable by a

11 word recognition engine for recognizing words." But if the second set of speech data values

12 contain sufficient information to recognize speech, and that second set of speech data values was

13 derived from the first set of speech data values, then the first set of speech data values logically

14 contained sufficient information to recognize the speech. *See* Rabiner Decl. ¶ 22.

15   Here, too, Phoenix responds by arguing that the claim is addressed to the further

16 processing of the first set of speech data values. That argument fails for the same reason.

17 Phoenix also responds that the second set of speech data values might be based on something

18 else in addition to the first set of speech data values. But nothing in the claim or the specification

19 or Phoenix's brief provides any clue as to what that mysterious additional information might be.

20 There is simply no support in the specification for the idea that the second set of speech data

21 values is derived from anything other than the first set of speech data values, as discussed in

22 more detail below, and thus no support in the specification for any such interpretation of the

23 claim.

24   d.   **The claims are indefinite because a person of skill in the art would not be**
             able to determine when speech data values are "insufficient" for speech
25   **recognition**

26   These claims are indefinite for an additional reason: they require a determination that the

27 speech data values created by the first signal processing circuit are "insufficient by themselves

28 for permitting recognition of the words" in an utterance. The problem is that there is no clear

430233.02

1   test for when a set of speech data values – or the data content of speech data – is or is not

2   sufficient for speech recognition.

3   Speech recognition is the process of determining what a user most likely said. It is

4   inherently a probabilistic exercise that may be more or less accurate. As Dr. Lawrence Rabiner

5   explains, the accuracy of speech recognition depends upon on the type of feature set used, the

6   amount of data that the feature set can accommodate, and the amount of data that is actually

7   used. *See* Rabiner Decl. ¶¶ 23-25. To take the example of the MFCC coefficients described in

8   the specification, there are potentially many such coefficients. And while a speech recognition

9   engine may be configured to accept 12 coefficients as inputs, that doesn't mean that speech

10 recognition cannot happen if fewer than 12 coefficients are used. To the contrary, one can input

11 12, or 4, or 8 coefficients derived from the speech utterance signals. If one inputs 12

12 coefficients, the results are likely to be very good. If one inputs 8, the results are likely to be less

13 good. And if one inputs only 4 coefficients, the results will probably be very unreliable. But

14 there is no bright-line test to distinguish what is "sufficient" from what is not, without knowing

15 what level of accuracy is required. Is 99% accuracy good enough? 75%? 50%? 2%? One of

16 skill in the art trying to understand whether a particular system would infringe the claims would

17 have no way of determining if a particular set of speech data values were "sufficient" for the

18 recognition of speech without a clear articulation of the standard by which sufficiency would be

19 judged. *See, e.g.*, *Honeywell Int'l, Inc. v. ITC*, 341 F.3d 1332, 1340 (Fed. Cir. 2003) (claim

20 indefinite where one of ordinary skill in the art would not understand the proper test to employ in

21 determining infringement); *Halliburton Energy Servs., Inc. v. M-I LLC*, 514 F.3d 1244, 1251

22 (Fed. Cir. 2008) ("Even if a claim term's definition can be reduced to words, the claim is

23 indefinite if a person of ordinary skill in the art cannot translate the definition into meaningfully

24 precise claim scope.").[6]

25

26 [6] Phoenix argues that the claim phrase is not indefinite because the examiner stated that the
    minimum amount of information to perform speech recognition is a **full set** of extracted speech
27 features. But the examiner said no such thing: in discussing a separate patent (the Barclay
    reference), the examiner stated that "extracted speech features are a minimum amount of
28 information that is sufficient" to allow speech recognition to occur. Trojan Decl., Ex. 6 at 3. As
    Dr. Lawrence Rabiner explains, saying that a set of extracted speech features are the least

11

WELLS FARGO'S RESPONSIVE MARKMAN BRIEF
CASE NO. CV 08-0863 MHP

430233.02

2.   **Signal processing functions required to generate said first and second set of speech data values can be allocated between said first signal processing circuit and second signal processing circuit as needed based on computing resources available to said first and second signal processing circuits respectively ('977 claims 1, 6, 7, 10)**

Signal processing functions required to generate said recognized speech query can be allocated between a client platform and the server computing system as needed based on computing resources available to said client platform and server computing system respectively ('846 claim 17)

Wells Fargo proposed constructions:

second set of speech data values	acoustic coefficients that have been derived from the "first set of speech data values"
recognized speech query	the system's best estimate of what a user actually said when compared to the words that are found in the system's dictionary file
allocated between … as needed based on resources available	signal processing functions required to derive "speech data values" can be done on both the client platform or server computing system, and those signal processing functions can be passed back and forth between client and server signal processing circuits depending on the computational capacity of both the client and server at a particular time

Phoenix proposed constructions:

entire claim phrases	The allocation of speech related functions performed by the client/server is configured based on computing resources available to such systems, meaning that the number of speech operations to be performed on an utterance by each is determined automatically (or with the assistance of a computer) by examining what resources are available for a recognition process at each. This allows for differing numbers of speech recognition operations to be done on the same utterance at the client/or server depending on resource availability

  a.   **Second set of speech data values**

The patent specification is clear that the second set of speech data values comprises

additional acoustic coefficients that are generated from the first set of speech data values. In the

particular embodiment described in the specification, the second set of speech data values are the

delta and acceleration coefficients, which are calculated from the 12 primary MFCC coefficients

and the energy coefficient. Trojan Decl., Ex. 1 at 16:22-26. But nowhere in the patent is there

---

information that can be used for speech recognition tells a person of skill in the art nothing about *how many* of the features must be used for speech recognition to occur. *See* Rabiner Decl. ¶ 26. And Phoenix's argument that a set of 12 MFCC vectors is insufficient to recognize speech is inconsistent with the specification, which suggests that 12 MFCC coefficients **sufficient** – not insufficient – for adequate and quick speech recognition in the preferred embodiment. Trojan Decl., Ex. 1 at 16:16-19 (MFCC parameters "represent the least amount of information that can be used by a subsequent server side system to adequately and quickly complete the recognition process").

12

430233.02

1 any disclosure of any source of information that could be used to generate the second set of

2 speech data values other than the first set of speech data values.

3   Phoenix refers to claim 32 of the '846 patent in an attempt to bolster its proposed

4 construction of "second set of speech data values" as encompassing things other than acoustic

5 coefficients, including "text words." But words are the outcome of the speech recognition

6 process, not something that "occurs during speech recognition." Unsurprisingly, then, nothing in

7 the specification or the claims suggests that text *words* are speech data *values*. Claim 32 depends

8 from claim 1 and 10, and neither of those non-asserted claims use the term "second set of speech

9 data values." Instead, the "second processing routine" is used to create "additional data content"

10 -- not "speech data values" -- that in the case of claim 32 is used to convert speech symbols into

11 text words.

12   **b.   Recognized speech query**

13   Phoenix argues that Wells Fargo's proposed definition is "unrelated to the claim term,"

14 but it doesn't say what else a recognized speech query might be, or what is wrong with Wells

15 Fargo's proposed construction. Br. at 6:14-7:3. Dr. Rabiner confirms that Wells Fargo's

16 proposed construction is correct. Rabiner Decl. ¶ 17.

17   **c.   Allocated between … as needed based on resources available**

18   These claims require that the signal processing functions required to generate the first and

19 second set of speech data values can be "allocated between" the client and the server – in other

20 words, that those functions can be performed by both the server and the client. These claims also

21 require that where the function is performed depends on "the computing resources available" to

22 each. In other words, this claim limitation is about the ability assign a particular function to

23 either the client or the server depending on the available resources. The patent specification

24 describes that, depending on how "thin" the client is, the derivation of the second set of speech

25 data values – the delta and acceleration coefficients – can take place at either the client or the

26 server. Trojan Decl., Ex. 1 at 22:40-46 ("In some environments, nonetheless, it is conceivable

27 that the MFCC delta parameters and MFCC acceleration parameters can also be computer at

28 client side system 150, depending on the computation resources available, the transmission

13

430233.02

1   bandwidth in data link 160A available to server side system 180, the speed of a transceiver used

2   for carrying data in the data link, etc.").

3   Phoenix's proposed construction is wrong because it could literally be read to cover any

4   situation where a client can perform or not perform a particular operation, regardless of whether

5   that operation could ever happen at the server. Under those circumstances, a different number of

6   speech recognition operations could be done at the client, but the different number of operations

7   would not result from an allocation between the client and the server. Instead, it would result

8   from a decision whether to perform the function *at all*. The same is true for a system that simply

9   uses a larger number of processing steps at the server when it has additional computational

10 resources to spare. That is not an "allocat[ion] **between**" client and server; at best it is an

11 allocation of certain functions **to** either the client or the server. That is not what the claim

12 requires.

13   Phoenix's proposed construction also uses the phrase "speech related functions," but the

14 claim language (like the specification) is clear that the functions to be allocated are not just any

15 speech related functions, but instead are limited to the "signal processing functions required to

16 generate speech data values" (or a recognized speech query). This distinction is important,

17 because Phoenix wants to argue that the claim limitation is met if *any* function related to speech

18 recognition can be allocated between the client and server.[7] That is not what is required by the

19 plain language of the claim, which is addressed to the allocation of those **signal processing**

20 **functions required** to generate the first and second set of speech data values. The first and

21 second set of speech data values are the speech vectors (acoustic coefficients); the claim thus

22 requires the ability to allocate the signal processing functions such that certain speech vectors

23 can be derived at either the client or the server. *See* Rabiner Decl. ¶ 15.

24

25

26 [7] Phoenix's infringement contentions accuse Wells Fargo of infringing this claim because the
    pruning of words may depend upon capacity at the server. In response to Wells Fargo's
    invalidity motion, however, Phoenix argues that the claim only covers the pruning of a statistical
27 language model grammar, not a manually written grammar. Whatever this supposedly means –
    and it's far from clear – nothing about this claim relates to kind of grammar being pruned, and
28 Phoenix does not contend otherwise in its claim construction briefing.

14

430233.02

3.   An answer file for storing a plurality of topic answer entries, each topic answer entry including an answer to one or more of said plurality of topic query entries, such that each topic query entry has at least one associated topic answer entry ('640 claims 1, 3, 4, 7)

Wells Fargo proposed constructions:

topic answer entries	files found in a database[8] that stores answers to specific questions. The answers are returned because they are connected to the query or queries, and not through a logical or deductive process. Topic answer entries are an answer to a particular query or queries. The topic answer entry is linked to that specific query or queries.

Phoenix proposed constructions:

entire claim phrase	one or more electronic files that store answers to topics covered by the IVR as entries in a table, database, other structured storage file. Each answer entry relates to one more topics which the user can ask questions about, and are arranged so that an answer can be given to any topic addressed by the IVR

The plain language of the claim requires that a topic answer entry include an answer to one or more topic query entries, and that each topic query entry have at least one associated topic answer entry. This means that the topic query entries and topic answer entries are associated with each other -- any user who asks a question that is matched to a given topic query entry will get the same topic answer entry, or entries, in return.[9] This is the process described in detail in the specification – (1) the Natural Language Engine matches the recognized words spoken by user to the closest question stored in the database, (2) that question is paired with an answer, and (3) that answer is returned as the "best answer" to the user's spoken query.

Phoenix tries to broaden this claim element well beyond its ordinary meaning, arguing that it will be satisfied whenever a system "can pick an answer based on a query." Br. at 9:27-10:1. That interpretation does away with the fixed relationship between topic query entries and topic answer entries. Wells Fargo's proposed construction that the answer not be reached through

---

[8] It is not clear whether Phoenix is trying to suggest that there is some distinction between tables and structured storage files on the one hand, and a "database" on the other. If so, it is also not clear what that distinction would be.

[9] The fact that there is more than one topic answer entry associated with a given query does not mean that the query has more than one answer. For example, topic answer entries could be "The Old Man and the Sea," "The Sun Also Rises," and "For Whom the Bell Tolls." The topic query entry "What books did Hemingway write?" would have all three of those topic answer entries associated with it. On the other hand, the topic query entry "what book did Hemingway write in 1940" would have only one of those topic answer entries -- "For Whom the Bell Tolls" -- associated with it. In either case, the topic query entry has a fixed answer, consisting of one or

15

430233.02

1 a logical or deductive process is simply intended to make clear that the answer is returned

2 because it is already "associated" with the query to which the user's request is matched, as the

3 claim language itself requires.

4 **4.   A database of query/answer pairs concerning one or more topics which can be responded to by the natural language query system during said interactive speech based session with a user (and similar claim phrases) ('854 claims 1, 7, 8, 9, 13, 15, 19, 20, 22, 23, 27, 28, 29)**

5

6   Wells Fargo proposed construction:

7

query/answer pairs	files found in a database that stores answers to specific questions. The answers are returned because they are connected to the query or queries, and not through a logical or deductive process. "Query/answer pairs" have a single answer linked to each question.

8

9

10   Phoenix proposed construction:

entire claim phrase	one or more electronic files that store answers to corresponding queries covered by the IVR in an organized and structured table. Each answer entry relates to one more queries which the user can ask with the IVR. Each of these stored questions has a paired answer stored in a separate text file, whose path is stored in a table of the database

11

12

13   Here, Phoenix appears to concede that the relationship of question to answer is fixed,

14 since each "question[] has a paired answer" under Phoenix's proposed definition. It is thus not

15 clear what it is in Wells Fargo's proposed construction to which Phoenix objects.

16 **5.   A natural language engine, which processes said recognized speech utterance data using a morphological analysis and a phrase analysis to form recognized speech sentence data corresponding to said speech-based query ('640 claims 1, 3, 4, 7)**

17

18 **A natural language routine adapted to process said first set of words and/or phrase and identify a response to said natural language query based on said query/answer pairs (and similar claim phrases) ('854 claims 1, 7, 8, 9, 13, 15, 19, 20, 22, 23, 23, 27, 28, 29)**

19

20

21   Wells Fargo proposed constructions:

"natural language engine" or "natural language routine"	operate on the recognized text that has been output by a speech recognition engine using morphological analysis and phrase analysis to interpret the words found in the recognized text.
morphological analysis	consists of determining the possible parts of speech for a particular word or group of words, *e.g.*, a noun phrase or a verb phrase.
phrase analysis	consists of determining the most likely meaning of a word or groups of words in recognized text by assigning parts of speech to particular words or groups of words based on the manner in which those words come together to form a sentence.

22

23

24

25

26

27

28 more topic answer entries, that it returns.

16

430233.02

Phoenix proposed constructions:

'640 patent entire claim phrase	A computer routine (in hardware or software form, or some combination thereof) that examines the recognized words (generated by a speech recognizer from a speech utterance) and analyzes them - as well as other forms of such recognized words - to identify selected sentences
'854 patent entire claim phrase	a natural language (NL) routine is a computer executed program that determines the meaning of a speech utterance. In this instance the NL route evaluates certain words and/or phrases presented by the speaker and determines the meaning of the query so that an appropriate response can be identified

Wells Fargo's proposed constructions follow directly from the usage of these terms in the specification. The specification describes the "basic linguistic morphological operations" as "tokenization, tagging and grouping." Trojan Decl., Ex. 1 at 17:50-51. Collectively, these three operations try to identify the part of speech for a particular word. As the specification explains, "tokenization … treats the text as a series of tokens or useful meaningful units that are larger than individual characters, but smaller than phrases and sentences." *Id*. at 17:53-56. Then "in the next phase of analysis, the tagger uses a built-in morphological analyzer to look up each word/token in a phrase or sentence and internally list all **parts of speech**." *Id*. at 17:62-65. "The output of the tagger is a string with each token tagged with **a parts of speech label**." *Id*. at 34:46-47. The tagger "has a built-in morphological analyzer … that allows it to identify **all possible parts of speech** for each token." *Id*. at 34:44-46. Then, finally, the grouper "determines which groups of words form phrases." *Id*. at 17:67-18:2. The specification explains that "**these three operations … are the foundations for any modern linguistic processing schemes**." *Id*. at 18:2-3. *See* Rabiner Decl. ¶ 28. The inventor of the patents also agrees with Wells Fargo's definition. Paige Decl., Ex. 3 at 169:16-21.

Phrase analysis can be understood as a subpart of this morphological analysis that is specifically directed to the contextual relationship between different words in a phrase. The specification explains that "a grouper routine 806 recognizes groups of words as phrases of a certain syntactic type," and that the function performed by the grouper is "very dependent, of course, on the performance and output of tagger component 804." Trojan Decl., Ex. 1 at 34:38-39, 38:50-51. *See* Rabiner Decl. ¶ 29. The tagger, in turn, "is a parts-of-speech disambiguator, which **analyzes words in context.**" Trojan Decl., Ex. 1 at 34:42-43. Indeed, in the prosecution

17

1   of related United States Patent No. 6,615,172, Phoenix explained that "linguistic processing is

2   distinct from plain word recognition, and involves concepts such as phrase analysis to fully

3   comprehend an entire closed question posed by a user, not merely picking out words to be used

4   in finding documents." Paige Decl., Ex. 8 ('172 Amendment A) at 19.[10]

5   Phoenix's proposed construction of this claim element suffers from several problems.

6   *First*, Phoenix jettisons the claim terms "morphological analysis" and "phrase analysis."

7   According to Phoenix, recognized words must merely be "analyzed" in some undefined way.

8   *Second*, Phoenix provides that the natural language engine operates "to identify selected

9   sentences", but does not give a hint as to what that means.

10   *Third*, there is no support in the specification for Phoenix's suggestion that the

11 recognized speech could come from somewhere other than the speech recognition engine, and

12 neither Phoenix nor the patent explains what source that might be.[11]

13   Wells Fargo proposes that the corresponding limitation in the '854 patent be interpreted

14 in the same way. Phoenix protests that the phrases "morphological analysis" and "phrase

15 analysis" don't expressly appear there. But the specification defines what it means by "natural

16 language processing" as including morphological analysis, and phrase analysis is encompassed

17 within it.

18   Here, too, Phoenix's proposed definition is of no use to a jury: it requires only that the

19 natural language routine "evaluate[] certain words and/or phrases" in an undefined manner so

20

[10] Although this statement comes from the prosecution history of the '172 patent, it may
nonetheless be used to interpret the meaning of the '640 patent's claim terms. *See Microsoft
Corp. v. Multi-Tech Sys., Inc.*, 357 F.3d 1340, 1350 (Fed. Cir. 2004) (explaining that
representations as to the nature of the invention from sibling applications can be used to interpret
claims when the patents share the same specification and the statement represents the applicant's
"representation of its own understanding of the invention disclosed in all three patents").

[11] The natural language engine must operate on "recognized speech utterance data" that is output
by the speech recognition engine. Indeed, in opposing Wells Fargo's summary judgment motion
of invalidity, Phoenix makes much of the point that the speech recognition engine and the natural
language engine are two distinct components, and that the latter operates on the output of the
former. In its summary judgment opposition, Phoenix explained that the patents-in-suit
described a system whereby speech was "first recognized, converted to words in the grammar,
then sent to the NLE to determine certain noun phrases." SJ Opp. at 22:2-3. Accordingly, in its
attempt to avoid summary judgment, Phoenix was clear that the two were separate components,
and that the natural language engine would obtain the data upon which it operated from the
speech recognition engine.

430233.02

WELLS FARGO'S RESPONSIVE MARKMAN BRIEF
CASE NO. CV 08-0863 MHP

1   that the "meaning of the query" is found and an "appropriate response" identified. This covers

2   the gamut of everything one could do with a speech-enabled IVR. In other words, an IVR that

3   merely had the phrase "press one for account balance" turned into "say one for account balance"

4   would arguably be covered by this definition, because once the IVR determined the user to have

5   said "one," the meaning of the query would be "give me my account balance," and the provision

6   of that balance would be the "appropriate response" to the query. That process is not the natural

7   language processing that Phoenix touted in the specification as part of the novelty of its

8   invention.

9   **6.   Is adapted to consider only a subset of said first set or words and/or phrases, and
10   further can consider words and/or phrases in said natural language query which are
             not present in said query/answer pairs to determine said response ('854 claim 1)**

11   Wells Fargo proposed construction:

entire claim phrase	evaluates words or phrases through the use of a "natural language engine," and is able to use that understanding to determine a response, even when the words used to obtain that response are not in the "query/answer pairs" found in the system

             Phoenix proposed construction:

entire claim phrase	the natural language routine only examines selected words and/or phrases presented by the speaker; moreover, the NL routine can identify a response even when the user presents words/phrases that are not part of the NL routine's grammar

17   Wells Fargo's proposed construction simply tracks the plain language of the claim.

18   Phoenix proposes that this phrase means that the routine can identify a response when the user's

19   query has words or phrases that are not part of the grammar. Phoenix's proposed construction

20   ignores the claim language in two key respects. First, the claim language requires that routine be

21   able to "<u>consider</u>" words and phrases in a query not found in the query/answer pairs, not that it

22   be able to <u>ignore</u> them. Second, the claim language requires that the system be able to consider

23   words present in the query which are not present in the <u>query/answer pairs</u>, not that it be able to

24   consider words not present in the system's <u>grammar</u>. Phoenix's proposed construction also does

25   not contain the requirements it maintained were a necessary part of this claim element in its

26   opposition to Wells Fargo's motion for summary judgment. There, Phoenix claimed that in

27   order to infringe (or anticipate), the natural language engine would have to have a grammar that

28   was "not entirely manually written," and was instead "automatically gleaned and used to

19

1   automatically train the system." SJ Opp. at 22:14-15. There is not a hint of this advanced

2   capability to be found in Phoenix's proposed construction. Wells Fargo's proposed construction,

3   on the other hand, properly focuses on both the query/answer pair aspect of the claim language,

4   and the fact that a system covered by the claim must consider words that are not part of the

5   query/answer pairs, not ignore those words and return a result regardless.

6   **7.   Said speech recognition engine is distributed between a client device and a server**
        system which receives streaming speech data having reduced latency data content
7   **before silence is detected and the utterance is complete ('854 claims 9, 27)**

8   Wells Fargo proposed construction:

entire claim phrase	a distributed voice recognition system divides speech recognition functions between a client and a server; both the client and the server create audio coefficients used for speech recognition
	the "speech data values" must be transmitted continuously rather than waiting for the end of the utterance and then sending "speech data values"

13   Phoenix proposed construction:

entire claim phrase	some speech recognition operations are performed on a client device which is connected by a network using an Internet protocol to a separate server system where other parts of the speech recognition process are performed. The speech data is sent continuously between the two systems only during times when the user is actually speaking, and until he/she is finished

17   Here, as in others of its remaining constructions, Phoenix reads in the requirement that

18   the system use an "Internet protocol." Phoenix does not explain where this requirement comes

19   from, nor point to any language in the specification limiting the invention to the use of Internet

20   protocol. Nor could it. The specification is notably broad when it comes to what could be used

21   to transmit data between portions of the system. *See* Trojan Decl., Ex. 1 at 10:54:57 ("Similarly,

22   while an INTERNET connection is depicted for data link 160A, it is apparent that **any channel**

23   that is suitable for carrying data between client system 150 and server system 180 will

24   suffice…."). The only other substantive disagreement between the parties here relates to what it

25   means to be "distributed," addressed in Part 11, *infra*.

26   **8.   A transmission circuit for formatting and transmitting said first set of speech data**
        values over a communications channel ('846, claim 11)

27   Wells Fargo proposed construction:

28

20

430233.02

entire claim phrase	a circuit that formats the "first set of speech data values," and then transmits the formatted "speech data values" over a link connecting the client and server device

Phoenix proposed construction:

entire claim phrase	a combination of hardware and software that prepares the speech data so that it can be sent in accordance with an Internet based protocol used by the communications channel

Here, too, the main substantive disagreement between the parties is that Phoenix reads

the claim to require the use of Internet protocol (a requirement with which the named inventor of

the patent himself disagrees, *see* Paige Decl., Ex. 3 at 229:17-22). In addition, Phoenix turns

"speech data values" into "speech data"; the former is required by the plain language of the

claims. Phoenix protests that Wells Fargo's definition requires those values to be first formatted

and then transmitted, but does not explain how any other temporal sequence would make sense:

how could the circuit format the values after having already transmitted them?

9.   Said first set of speech data values are sent in a streaming fashion over said channel before silence is detected and/or said speech utterance is completed (846, claim 11)

Wells Fargo proposed construction:

entire claim phrase	the "speech data values" must be transmitted continuously rather than waiting for the end of the utterance and then sending "speech data values"

Phoenix proposed construction:

entire claim phrase	the system can operate such that from the moment the user starts speaking the captured speech data is sent continuously over a channel coupling the user to the second signal processing circuit; the data is sent only during times when there is voice activity, and/or until he/she is finished speaking

Wells Fargo's construction tracks the plain meaning of the claim and is supported by the

specification, which distinguishes over the prior art on the basis that "the streaming of acoustic

parameters does not appear to be implemented in real time as it can only take place after silence

is detected." Trojan Decl., Ex. 1 at 5:55-57. In its proposed definition, Phoenix once more uses

the vague term "speech data" rather than the "speech data values" required by the claim.

Moreover, the channel does not link the user to the second signal processing circuit, as Phoenix

provides in its construction. Instead, it couples the first signal processing circuit (which

produces speech data values) to the second signal processing circuit.

21

430233.02

10.   Remote speech capturing system ('640, claim 1)

Wells Fargo proposed construction:

remote speech capturing system	a client located in another physical location from the server that has the ability to perform speech recognition operations on the speech signals that it receives from a user

Phoenix proposed construction:

remote speech capturing system	the speech data is captured by a computing system that is physically separate from the speech recognition system but coupled to the latter by a network using an Internet based protocol

The only substantive disagreement between the parties appears to be whether the claim requires the use of a Internet protocol. For the reasons set forth above, it does not.[12]

11.   Distributed voice recognition system ('846, claims 11, 17, 20, 21)/distributed ('854, claims 9, 23, 27, 28, 29)

Wells Fargo proposed construction:

distributed voice recognition system/distributed	divides speech recognition functions between a client and a server; both the client and the server create audio coefficients used for speech recognition [or process those audio coefficients into recognized text]

Phoenix proposed construction:

distributed voice recognition system	some speech recognition operations are performed on a client device which is connected by a network using an Internet protocol to a separate server system where other parts of the speech recognition process are performed
distributed	some tasks are performed on a first device which is connected by a network using an Internet protocol to a separate device

Here, too, Phoenix tries to read in the requirement that the system use an Internet protocol. In addition to that problem, Phoenix conjures up the requirement that the client device be connected to the server by "a network" without explaining what constitutes a network or where that limitation appears in the claim.

Wells Fargo's construction is intended to reflect the idea, described throughout the specification, that the speech recognition functions are distributed between the client and the server. In that regard, it is important to recognize that what is "distributed" in these claims is the

---

[12] Notably, however, Phoenix's proposed construction also appears to contradict its arguments in opposition to Wells Fargo's motion for summary judgment. There, Phoenix took the position that a remote speech capturing system within the meaning of the '640 patent was "one where a *client* is connected to a server through an communications channel," SJ Opp. at 19:4-6 (emphasis added). Now, however, Phoenix appears to disagree that the remote speech capturing system is a client located at a physical location remote from the server.

WELLS FARGO'S RESPONSIVE MARKMAN BRIEF
CASE NO. CV 08-0863 MHP

430233.02

1   "speech recognition engine" (in the '854 patent) or the "voice recognition system" that is

2   comprised of the "first signal processing circuit" and the "second signal processing circuit" (in

3   the '846 patent) which create the first and second set of speech vectors.[13]

4   **12.   A speech-enabled internet website operating on a server computing system ('977, claims 1, 6, 7, 10)**

5   Wells Fargo proposed construction:

6

7 | speech-enabled internet website | a group of related pages on the World Wide Web that are visible to and navigable by an end user using a web browser through spoken commands |
     | --- | --- |

8   Phoenix proposed construction:

9 | speech-enabled internet website operating on a server computing system | a server computing system with speech capability and which has at least one internet server having one more web pages that can be interpreted by a browser |
     | --- | --- |

10

11   Once again, Phoenix seeks to ignore an important element of the claimed invention.

12   Phoenix contends, without any citation to the specification or claim, that the preamble phrase

13   "speech-enabled internet website" is not a claim limitation because it "merely sets out the

14   functionalities and attributes already explained in the body of the claim." Br. at 17:23-24.

15   Phoenix is mistaken for two independent reasons.

16   *First*, the phrase "speech-enabled internet website" identifies a fundamental aspect of the

17   invention: a way to interact verbally with the pages of websites instead of having to click on

18   them with a mouse. The specification explains that "[t]he invention relates to a system and an

19   interactive method for an [sic] enabling a website to have interactive, real-time speech-enabled

20   web pages … so that a user can intelligently and easily control an internet session using a

21   conventional browser that is enhanced to handle speech capabilities." Trojan Decl., Ex. 3 ('977

22   patent) at 1:19-26. It describes the problems associated with then-existing ways of interacting

23   with internet websites: "the INTERNET 'experience' for users has been limited mostly to non-

24   voice based input/output devices" but that "many persons cannot or will not, because of physical

25

26   _____

     [13] Phoenix takes issue with the proposition that the creation of audio coefficients must be
     distributed between the server and the client, and refers to disclosures in the specification that, in

27   an alternate embodiment, all feature extraction could be performed at the client. In that
     circumstance, the "speech recognition engine" would be distributed in the sense that all the
     feature vectors would be generated at the client, but the recognized text would be generated at

28   the server. Fair point. We have amended our proposed construction accordingly.

─────────────────────────────────────

23

430233.02

1   or psychological barriers, use any of the aforementioned conventional … devices." *Id*. at 1:48-

2   51, 1:63-66. The specification then explains that the invention is explicitly directed at solving

3   these problems: "the present invention is to provide an INTERNET website with speech

4   processing capability so that voice based data and commands can be used to interact with such

5   site. . . ." *Id*. at 6:24-27. Because a speech-enabled internet website is fundamental to the

6   claimed invention, the corresponding preamble phrase necessarily limits the claims. *See Poly-*

7   *America, L.P. v. GSE Lining Tech., Inc.*, 383 F.3d 1303, 1310 (Fed. Cir. 2004) (preamble

8   "discloses a fundamental characteristic of the claimed invention" and thus "is properly construed

9   as a limitation of the claim itself"); *On Demand Mach. Corp. v. Ingram Indus*., 442 F.3d 1331,

10   1344 (Fed. Cir. 2006) (preamble requiring high speed manufacture of a single copy was a

11   limitation because "the high speed manufacture of a single copy is fundamental to the Ross

12   invention.").

13   ***Second***, the preamble was relied upon to distinguish the claimed invention from the prior

14   art. In the prosecution of the '977 patent, the examiner explained the reasons for allowance:

15   [T]he prior art record does not disclose or reasonably suggest the combination of
             a ***speech-enabled website***, where a server receives speech data associated with a
16   user query, where data content is substantially inadequate by itself for permitting
             recognition of words or involves limited speech data content to reduce processing
17   and transmission latencies, a web page with a list of items or with a search engine,
             and wherein signal processing functions are allocated between a client and a
18   server as needed based on available computing resources.

19   Paige Decl., Ex. 9 ('977 Reasons For Allowance) at 2 (emphasis added). Such clear reliance on

20   the "speech-enabled website" preamble phrase to distinguish prior art "transforms the preamble

21   into a claim limitation because such reliance indicates use of the preamble to define, in part, the

22   claimed invention." *Catalina Mktg. Int'l v. Coolsavings.com*, *Inc.*, 289 F.3d 801, 808 (Fed. Cir.

23   2002).

24   Moreover, Phoenix's proposed construction of "speech-enabled website" would also read

25   that requirement out of the claims. Phoenix contends that a website is "speech-enabled" if the

26   information on the website can be separately accessed using any speech-enabled system—even

27   one that is entirely separate from the website. For example, Phoenix argues that to the extent

28   that Lands' End has a website that offers products for sale and it is possible to order those same

430233.02

1   products from the Lands' End catalogue over the phone using a speech-enabled system, Lands'

2   End has a "speech-enabled" website. That has nothing to do with the invention, which allows "a

3   user [to] intelligently and easily control an internet session using a conventional browser that is

4   enhanced to handle speech capabilities." Trojan Decl., Ex. 3 at 1:19-21; Paige Decl., Ex. 3 at

5   218:17-220:2 (speech enabled website accepts voice input).

6   Phoenix also takes issue with the proposition that a website is visible to a user. But

7   people use browsers to explore the world wide web visually; even websites designed for the

8   blind have a visual component. *See, e.g.*, http://fonetiks.org/websitesforblindpeople.html.[14]

9   Phoenix also argues that "internet website" could in fact mean an intranet website or other

10   network. But the specification distinguishes between the internet and a local intranet Trojan

11   Decl., Ex. 3 at 6:59-60, and Phoenix chose to use the term "internet website" in defining its

12   claims. Phoenix cannot now abandon that aspect of the claim.

13   Finally, Phoenix relies on the doctrine of claim differentiation, arguing that the speech-

14   enabled website of claim 1 must not be "navigable by an end user" because claim 3 adds the

15   requirement that the website be adapted such that a user can "navigate and locate information of

16   interest" by way of speech query. But claim 3 introduces an additional concept—the ability not

17   just to navigate but to locate information of interest using the speech query. In any event, "the

18   presumption created by the doctrine of claim differentiation is not a hard and fast rule and will be

19   overcome by a contrary construction dictated by the written description or prosecution history,"

20   and any inconclusive inference drawn from claim 3 cannot overcome the weight of the intrinsic

21   evidence supporting Wells Fargo's construction. *Regents of the Univ. of Cal. v.*

22   *DakoCytomation Cal., Inc.*, 517 F.3d 1364, 1375 (Fed. Cir. 2008).

23   **IV.   CONCLUSION**

24   For the foregoing reasons, Wells Fargo respectfully requests that the Court adopt its

25   proposed claim constructions.

26   ───────────────────

27   [14] Indeed, Phoenix's citation of claim 33 in its claim differentiation argument supports Wells
     Fargo's point. It requires a website that is able to accept input through a pointing device or
     keyboard. A website that can accept input through a pointing device must have a visual
28   component in order to allow the use of the pointing device.

430233.02   WELLS FARGO'S RESPONSIVE MARKMAN BRIEF
                            CASE NO. CV 08-0863 MHP

1

Dated: November 13, 2008   KEKER & VAN NEST, LLP

2

3

4   By: /s/ Eugene M. Paige_____
                                                      EUGENE M. PAIGE
5   Attorneys for Defendants
                                                      WELLS FARGO BANK, N.A. and
6   WELLS FARGO FUNDS
                                                      MANAGEMENT, LLC
7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

26

430233.02